

# Super-Exponential Solution in Markovian Supermarket Models: Framework and Challenge

Quan-Lin Li

School of Economics and Management Sciences  
Yanshan University, Qinhuangdao 066004, China

January 28, 2011

## Abstract

Marcel F. Neuts opened a key door in numerical computation of stochastic models by means of phase-type (PH) distributions and Markovian arrival processes (MAPs). To celebrate his 75th birthday, this paper reports a more general framework of Markovian supermarket models, including a system of differential equations for the fraction measure and a system of nonlinear equations for the fixed point. To understand this framework heuristically, this paper gives a detailed analysis for three important supermarket examples: M/G/1 type, GI/M/1 type and multiple choices, explains how to derive the system of differential equations by means of density-dependent jump Markov processes, and shows that the fixed point may be simply super-exponential through solving the system of nonlinear equations. Note that supermarket models are a class of complicated queueing systems and their analysis can not apply popular queueing theory, it is necessary in the study of supermarket models to summarize such a more general framework which enables us to focus on important research issues. On this line, this paper develops matrix-analytical methods of Markovian supermarket models. We hope this will be able to open a new avenue in performance evaluation of supermarket models by means of matrix-analytical methods.

**Keywords:** Randomized load balancing, supermarket model, matrix-analytic method, super-exponential solution, density-dependent jump Markov process, Batch Markovian Arrival Process (BMAP), phase-type (PH) distribution, fixed point.

# 1 Introduction

In the study of Markovian supermarket models, this paper proposes a more general framework including a system of differential equations for the fraction measure and a system of nonlinear equations for the fixed point, and the both systems of equations enable us to focus on important research issues of Markovian supermarket models. At the same time, this paper indicates that it is difficult and challenging to analyze the system of differential equations and to solve the system of nonlinear equations from four key directions: Existence of solution, uniqueness of solution, stability of solution and effective algorithms. Since there is a large gap to provide a complete solution to the both systems of equations, this paper devotes heuristic understanding of how to organize and solve the both systems of equations by means of discussing three important supermarket examples: M/G/1 type, GI/M/1 type and multiple choices. Specifically, the supermarket examples show a key result that the fixed point can be super-exponential for more supermarket models. Note that supermarket models are a class of complicated queueing systems and their analysis can not apply popular queueing theory, while recent research gave some simple and beautiful results for special supermarket models, e.g., see Mitzenmacher [19], Li and Lui [11] and Luczak and McDiarmid [14], this motivates us in this paper to summarize a more general framework in order to develop matrix-analytical methods of Markovian supermarket models. We hope this is able to open a new avenue for performance evaluation of supermarket models by means of matrix-analytical methods.

Recently, a number of companies, such as Amazon and Google, are offering cloud computing service and cloud manufacturing technology. This motivates us in this paper to study randomized load balancing for large-scale networks with many computational and manufacturing resources. Randomized load balancing, where a job is assigned to a server from a small subset of randomly chosen servers, is very simple to implement. It can surprisingly deliver better performance (for example reducing collisions, waiting times and backlogs) in a number of applications including data centers, distributed memory machines, path selection in computer networks, and task assignment at web servers. Supermarket models are extensively used to study randomized load balancing schemes. In the past ten years, supermarket models have been studied by queueing theory as well as Markov processes. Since the study of supermarket models can not apply popular queueing theory, they have not been extensively studied in queueing committee up to now. There-

fore, this leads to that available queueing results of supermarket models are few up to now. Some recent works dealt with the supermarket model with Poisson arrivals and exponential service times by means of density-dependent jump Markov processes, discussed limiting behavior of the supermarket model under a weakly convergent setting when the population size goes to infinite, and indicated that there exists a doubly exponential solution to the fixed point through solving the system of nonlinear equations. Readers may refer to population processes by Kurtz [8], and doubly exponential solution with exponential improvement by Vvedenskaya, Dobrushin and Karpelevich [27], Mitzenmacher [19], Li and Lui [11] and Luczak and McDiarmid [14].

Certain generalization of supermarket models has been explored in, for example, studying simple variations by Vvedenskaya and Suhov [28], Mitzenmacher [20], Azar, Broder, Karlin and Upfal [1], Vöcking [26], Mitzenmacher, Richa, and Sitaraman [22] and Li, Lui and Wang [13]; considering non-Poisson arrivals or non-exponential service times by Li, Lui and Wang [12], Li and Lui [11], Bramson, Lu and Prabhakar [2] and Li [10]; discussing load information by Mirchandaney, Towsley, and Stankovic [23], Dahlin [3] and Mitzenmacher [21]; mathematical analysis by Graham [4, 5, 6], Luczak and Norris [16] and Luczak and McDiarmid [14, 15]; using fast Jackson networks by Martin and Suhov [18], Martin [17] and Suhov and Vvedenskaya [25].

The main contributions of the paper are twofold. The first one is to propose a more general framework for Markovian supermarket models. This framework contains a system of differential equations for the fraction measure and a system of nonlinear equations for the fixed point. It is indicated that there exist more difficulties and challenges for dealing with the system of differential equations and for solving the system of nonlinear equations because of two key factors: infinite dimension and complicated structure of nonlinear equations. Since there is still a large gap up to being able to deal with the both systems of equations systematically, the second contribution of this paper is to analyze three important supermarket examples: M/G/1 type, GI/M/1 type and multiple choices. These examples provide necessary understanding and heuristic methods in order to discuss the both systems of equations from practical and more general point of view. For the supermarket examples, this paper derives the systems of differential equations for the fraction measure by means of density-dependent jump Markov processes, and illustrates that the fixed points may be super-exponential through solving the systems of nonlinear equations by means of matrix-analytic methods.

The remainder of this paper is organized as follows. Section 2 proposes a more general framework for Markovian supermarket models. This framework contains a system of differential equations for the fraction measure and a system of nonlinear equations for the fixed point. In Sections 3 and 4, we consider a supermarket model of M/G/1 type by means of BMAPs and a supermarket model of GI/M/1 type in terms of batch PH service processes, respectively. For the both supermarket models, we derive the systems of differential equations satisfied by the fraction measure in terms of density-dependent jump Markov processes, and obtain the system of nonlinear equations satisfied by the fixed point which is shown to be super-exponential. In Section 5, we analyze two supermarket models with multiple choice numbers, and give super-exponential solution to the fixed points for the two supermarket models. Note that the supermarket examples discussed in Sections 3 to 5 can provide a heuristic understanding for the more general framework of Markovian supermarket model given in Section 2.

## 2 Markovian Supermarket Models

In this section, we propose a more general framework for Markovian supermarket models. This framework contains a system of differential equations for the fraction measure and a system of nonlinear equations for the fixed point.

Recent research, e.g., see Mitzenmacher [20] and Li and Lui [11], shows that a Markovian supermarket model contains two important factors:

- (1) Continuous-time Markov chain  $Q$ , called stochastic environment of the supermarket model; and
- (2) Choice numbers, including input choice numbers  $d_1, d_2, \dots, d_v$  and output choice numbers  $f_1, f_2, \dots, f_w$ . Note that the choice numbers determine decomposed structure of the stochastic environment  $Q$ .

We first analyze stochastic environment of the Markovian supermarket model. From point of view of stochastic models, we take a more general stochastic environment which is a continuous-time Markov chain  $\{X_t, t \geq 0\}$  with block structure. We assume that the Markov chain  $\{X_t, t \geq 0\}$  on state space  $\Omega = \{(k, j) : k \geq 0, 1 \leq j \leq m_k\}$  is irreducible

and positive recurrent, and that its infinitesimal generator is given by

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & \cdots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,0} & Q_{2,1} & Q_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (1)$$

where  $Q_{i,j}$  is a matrix of size  $m_i \times m_j$  whose  $(r, r')$ th entry is the transition rate of the Markov chain from state  $(i, r)$  to state  $(j, r')$ . It is well-known that  $Q_{i,j} \geq 0$  for  $i \neq j$ ,  $Q_{i,i}$  is invertible with strictly negative diagonal entries and nonnegative off-diagonal entries. For state  $(i, k)$ ,  $i$  is called the *level variable* and  $k$  the *phase variable*. We write level  $i$  as  $L_i = \{(i, k) : 1 \leq k \leq m_i\}$ .

Since the Markov chain is irreducible, for each level  $i$  there must exist at east one left-block state transition:  $\leftarrow$  level  $i$  or level  $i \leftarrow$ , and at east one right-block state transition:  $\rightarrow$  level  $i$  or level  $i \rightarrow$ . We write

$$E_{\text{left}} = \{\leftarrow \text{ level } i \text{ or level } i \leftarrow : \text{ level } i \in \Omega\}$$

and

$$E_{\text{right}} = \{\rightarrow \text{ level } i \text{ or level } i \rightarrow : \text{ level } i \in \Omega\}.$$

Note that  $E_{\text{left}}$  and  $E_{\text{right}}$  describe output and input processes in the supermarket model. Based on the two block-transition sets  $E_{\text{left}}$  and  $E_{\text{right}}$ , we write

$$Q = Q_{\text{left}} + Q_{\text{right}}, \quad (2)$$

and for  $i \geq 0$

$$Q_{i,i} = Q_i^{\text{left}} + Q_i^{\text{right}}.$$

Thus we have

$$Q_{\text{left}} = \begin{pmatrix} Q_0^{\text{left}} & & & \\ Q_{1,0} & Q_1^{\text{left}} & & \\ Q_{2,0} & Q_{2,1} & Q_2^{\text{left}} & \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and

$$Q_{\text{right}} = \begin{pmatrix} Q_0^{\text{right}} & Q_{0,1} & Q_{0,2} & \cdots \\ & Q_1^{\text{right}} & Q_{1,2} & \cdots \\ & & Q_2^{\text{right}} & \cdots \\ & & & \ddots \end{pmatrix}.$$

Note that  $Qe = 0$ ,  $Q_{\text{left}}e = 0$  and  $Q_{\text{right}}e = 0$ , where  $e$  is a column vector of ones with a suitable dimension in the context. We assume that the matrices  $Q_j^{\text{left}}$  for  $j \geq 1$  and  $Q_i^{\text{right}}$  for  $i \geq 0$  are all invertible, while  $Q_0^{\text{left}}$  is possibly singular if there is not an output process in level 0. We call  $Q = Q_{\text{left}} + Q_{\text{right}}$  an input-output rate decomposition of the Markovian supermarket model.

Now, we provide a choice decomposition of the Markovian supermarket model through decomposing the two matrices  $Q_{\text{left}}$  and  $Q_{\text{right}}$ . Note that the choice decomposition is based on the input choice numbers  $d_1, d_2, \dots, d_v$  and the output choice numbers  $f_1, f_2, \dots, f_w$ . We write

$$Q_{\text{left}} = Q_{\text{left}}(f_1) + Q_{\text{left}}(f_2) + \dots + Q_{\text{left}}(f_w) \quad (3)$$

for the output choice numbers  $f_1, f_2, \dots, f_w$ , and

$$Q_{\text{right}} = Q_{\text{right}}(d_1) + Q_{\text{right}}(d_2) + \dots + Q_{\text{right}}(d_v) \quad (4)$$

for the input choice numbers  $d_1, d_2, \dots, d_v$ .

To study the Markovian supermarket model, we need to introduce two vector notation. For a vector  $a = (a_1, a_2, a_3, \dots)$ , we write

$$a^{\odot d} = \left( a_1^d, a_2^d, a_3^d, \dots \right)$$

and

$$a^{\odot \frac{1}{d}} = \left( a_1^{\frac{1}{d}}, a_2^{\frac{1}{d}}, a_3^{\frac{1}{d}}, \dots \right).$$

Let  $S(t) = (S_0(t), S_1(t), S_2(t), \dots)$  be the fraction measure of the Markovian supermarket model, where  $S_i(t)$  is a row vector of size  $m_i$  for  $i \geq 0$ . Then  $S(t) \geq 0$  and  $S_0(t)e = 1$ . Based on the input-output rate decomposition and the choice decomposition for the stochastic environment, we introduce the following system of differential equations satisfied by the fraction measure  $S(t)$  as follows:

$$S_0(t) \geq 0 \text{ and } S_0(t)e = 1, \quad (5)$$

and

$$\frac{d}{dt}S(t) = \sum_{l=1}^w S^{\odot f_l}(t) Q_{\text{left}}(f_l) + \sum_{k=1}^v S^{\odot d_k}(t) Q_{\text{right}}(d_k). \quad (6)$$

In the Markovian supermarket model, a row vector  $\pi = (\pi_0, \pi_1, \pi_2, \dots)$  is called a fixed point of the fraction measure  $S(t)$  if  $\lim_{t \rightarrow +\infty} S(t) = \pi$ . In this case, it is easy to see that

$$\lim_{t \rightarrow +\infty} \left[ \frac{d}{dt}S(t) \right] = 0.$$

If there exists a fixed point of the fraction measure, then it follows from (5) and (6) that the fixed point is a nonnegative non-zero solution to the following system of nonlinear equations

$$\pi_0 \geq 0 \text{ and } \pi_0 e = 1, \quad (7)$$

and

$$\sum_{l=1}^w \pi^{\odot f_l} Q_{\text{left}}(f_l) + \sum_{k=1}^v \pi^{\odot d_k} Q_{\text{right}}(d_k) = 0. \quad (8)$$

**Remark 1** *If  $d_k = 1$  for  $1 \leq k \leq v$  and  $f_l = 1$  for  $1 \leq l \leq w$ , then the system of differential equations (5) and (6) is given by*

$$S_0(t) \geq 0 \text{ and } S_0(t) e = 1,$$

and

$$\frac{d}{dt} S(t) = S(t) Q.$$

Thus we obtain

$$S(t) = c S(0) \exp \{Qt\}.$$

Let

$$W(t) = (W_0(t), W_1(t), W_2(t), \dots) = S(0) \exp \{Qt\},$$

where  $W_i(t)$  is a row vector of size  $m_i$  for  $i \geq 0$ . Then  $S(t) = cW(t)$ , where  $c = 1/W_0(t)e$ . At the same time, the system of nonlinear equations (7) and (8) is given by

$$\pi_0 \geq 0 \text{ and } \pi_0 e = 1,$$

and

$$\pi Q = 0.$$

Let  $W = (w_0, w_1, w_2, \dots)$  be the stationary probability vector of the Markov chain  $Q$ , where  $W_i$  is a row vector of size  $m_i$  for  $i \geq 0$ . Then  $\pi = cW$ , where  $c = 1/w_0 e$ . Note that the stationary probability vector  $W$  of the block-structured Markov chain  $Q$  is given a detailed analysis in Chapter 2 of Li [9] by means of the RG-factorizations.

If there exist some  $d_k \geq 2$  or/and  $f_l \geq 2$  in the Markovian supermarket model, then the system of differential equations (5) and (6) and the system of nonlinear equations (7) to (8) are two decomposed power-form generalizations of transient solution and of stationary probability of an irreducible continuous-time Markov chain with block structure

(see Chapters 2 and 8 of Li [9]). Note that Li [9] can deal with transient solution and stationary probability for an irreducible block-structured Markov chain, where the RG-factorizations play a key role. However, the RG-factorizations can not hold for Markovian supermarket models with some  $d_k \geq 2$  or/and  $f_l \geq 2$ . Therefore, there exist more difficulties and challenges to study the system of differential equations (5) and (6) and the system of nonlinear equations (7) to (8). Specifically, it still keeps not to be able to answer four important issues: Existence of solution, uniqueness of solution, stability of solution and effective algorithms. This is similar to some research on the four important issues of irreducible continuous-time Markov chains with block structure.

In the remainder of this paper, we will study three important Markovian supermarket examples: M/G/1 type, GI/M/1 type and multiple choices. Our purpose is to provide heuristic understanding of how to set up and solve the system of differential equations (5) and (6), and the system of nonlinear equations (7) to (8).

### 3 A Supermarket Model of M/G/1 Type

In this section, we consider a supermarket model with a BMAP and exponential service times. Note that the stochastic environment is a Markov chain of M/G/1 type, the supermarket model is called to be of M/G/1 type. For the supermarket model of M/G/1 type, we set up the system of differential equations for the fraction measure by means of density-dependent jump Markov processes, and derive the system of nonlinear equations satisfied by the fixed point which is shown to be super-exponential solution.

The supermarket model of M/G/1 type is described as follows. Customers arrive at a queueing system of  $n > 1$  servers as a BMAP with irreducible matrix descriptor  $(nC, nD_1, nD_2, nD_3, \dots)$  of size  $m$ , where the matrix  $C$  is invertible and has strictly negative diagonal entries and nonnegative off-diagonal;  $D_k \geq 0$  is the arrival rate matrix with batch size  $k$  for  $k \geq 1$ . We assume that  $\sum_{k=1}^{\infty} kD_k$  is finite and that  $C + \sum_{k=1}^{\infty} D_k$  is an irreducible infinitesimal generator with  $(C + \sum_{k=1}^{\infty} D_k)e = 0$ . Let  $\gamma$  be the stationary probability vector of the irreducible Markov chain  $C + \sum_{k=1}^{\infty} D_k$ . Then the stationary arrival rate of the BMAP is given by  $n\lambda = n\gamma \sum_{k=1}^{\infty} kD_k e$ . The service times of each customer are exponentially distributed with service rate  $\mu$ . Each batch of arriving customers choose  $d \geq 1$  servers independently and uniformly at random from the  $n$  servers, and joins for service at the server which currently possesses the fewest number of customers.



If there is a tie, servers with the fewest number of customers will be chosen randomly. All customers in every server will be served in the first-come-first service (FCFS) manner. We assume that all the random variables defined above are independent of each other and that this system is operating in the region  $\rho = \lambda/\mu < 1$ . Clearly,  $d$  is an input choice number in this supermarket model. Figure 1 is depicted as an illustration for supermarket models of M/G/1 type.

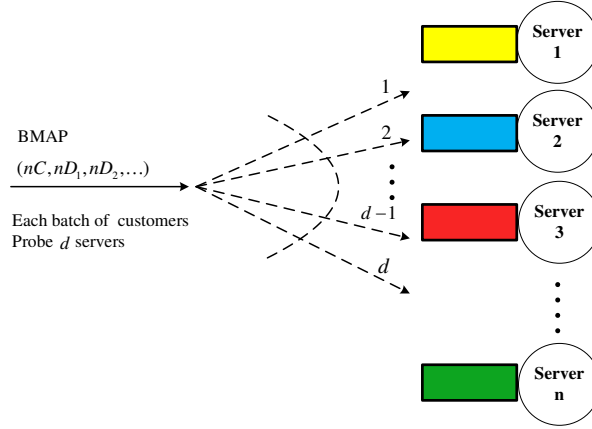


Figure 1: A supermarket model of M/G/1 type

The supermarket model with a BMAP and exponential service times is stable if  $\rho = \lambda/\mu < 1$ . This proof can be given by a simple comparison argument with the queueing system in which each customer queues at a random server (i.e., where  $d = 1$ ). When  $d = 1$ , each server acts like a BMAP/M/1 queue which is stable if  $\rho = \lambda/\mu < 1$ , see chapter 5 in Neuts [24]. Similar to analysis in Winston [30] and Weber [29], the comparison argument leads to two useful results: (1) the shortest queue is optimal due to the assumptions on a BMAP and exponential service times in the supermarket model; and (2) the size of the longest queue in the supermarket model is stochastically dominated by the size of the longest queue in a set of  $n$  independent BMAP/M/1 queues.

We define  $n_k^{(i)}(t)$  as the number of queues with at least  $k$  customers, including customers in service, and with the BMAP in phase  $i$  at time  $t \geq 0$ . Clearly,  $0 \leq n_k^{(i)}(t) \leq n$  for  $k \geq 0$  and  $1 \leq i \leq m$ . Let

$$x_n^{(i)}(k, t) = \frac{n_k^{(i)}(t)}{n},$$

which is the fraction of queues with at least  $k$  customers and the BMAP in phase  $i$  at

time  $t \geq 0$  for  $k \geq 0$ . We write

$$X_n(k, t) = \left( x_n^{(1)}(k, t), x_n^{(2)}(k, t), \dots, x_n^{(m)}(k, t) \right)$$

for  $k \geq 0$ , and

$$X_n(t) = (X_n(0, t), X_n(1, t), X_n(2, t), \dots).$$

The state of the supermarket model may be described by the vector  $X_n(t)$  for  $t \geq 0$ . Since the arrival process to the queueing system is a BMAP and the service time of each customer is exponential, the stochastic process  $\{X_n(t), t \geq 0\}$  is a Markov process whose state space is given by

$$\Omega_n = \{ (g_n^{(0)}, g_n^{(1)}, g_n^{(2)}, \dots) : g_n^{(0)} \text{ is a probability vector, } g_n^{(k)} \geq g_n^{(k+1)} \geq 0 \text{ for } k \geq 1, \text{ and } ng_n^{(l)} \text{ is a vector of nonnegative integers for } l \geq 0 \}.$$

Let

$$s_k^{(i)}(n, t) = E \left[ x_k^{(i)}(n, t) \right],$$

and

$$S_k(n, t) = \left( s_k^{(1)}(n, t), s_k^{(2)}(n, t), \dots, s_k^{(m)}(n, t) \right)$$

for  $k \geq 0$ ,

$$S(n, t) = (S_0(n, t), S_1(n, t), S_2(n, t), \dots).$$

As shown in Martin and Suhov [18] and Luczak and McDiarmid [14], the Markov process  $\{X_n(t), t \geq 0\}$  is asymptotically deterministic as  $n \rightarrow \infty$ . Thus  $\lim_{n \rightarrow \infty} E \left[ x_k^{(i)}(n, t) \right]$  always exist by means of the law of large numbers for  $k \geq 0$ . Based on this, we write

$$S_k(t) = \lim_{n \rightarrow \infty} S_k(n, t)$$

for  $k \geq 0$ , and

$$S(t) = (S_0(t), S_1(t), S_2(t), \dots).$$

Let  $X(t) = \lim_{n \rightarrow \infty} X_n(t)$ . Then it is easy to see from the BMAP and the exponential service times that  $\{X(t), t \geq 0\}$  is also a Markov process whose state space is given by

$$\Omega = \left\{ (g^{(0)}, g^{(1)}, g^{(2)}, \dots) : g^{(0)} \text{ is a probability vector, } g^{(k)} \geq g^{(k+1)} \geq 0 \text{ for } k \geq 1 \right\}.$$

If the initial distribution of the Markov process  $\{X_n(t), t \geq 0\}$  approaches the Dirac delta-measure concentrated at a point  $g \in \Omega$ , then  $X(t) = \lim_{n \rightarrow \infty} X_n(t)$  is concentrated on

the trajectory  $S_g = \{S(t) : t \geq 0\}$ . This indicates a law of large numbers for the time evolution of the fraction of queues of different lengths. Furthermore, the Markov process  $\{X_n(t), t \geq 0\}$  converges weakly to the fraction vector  $S(t) = (S_0(t), S_1(t), S_2(t), \dots)$  as  $n \rightarrow \infty$ , or for a sufficiently small  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P\{\|X_n(t) - S(t)\| \geq \varepsilon\} = 0,$$

where  $\|a\|$  is the  $L_\infty$ -norm of vector  $a$ .

In what follows we set up a system of differential vector equations satisfied by the fraction vector  $S(t)$  by means of density-dependent jump Markov processes.

We first provide an example to indicate how to derive the differential vector equations. Consider the supermarket model with  $n$  servers, and determine the expected change in the number of queues with at least  $k$  customers over a small time interval  $[0, dt)$ . The probability vector that an arriving customer joins a queue with  $k-1$  customers in this time interval is given by

$$\left[ S_0^{\odot d}(n, t) D_k + S_1^{\odot d}(n, t) D_{k-1} + \dots + S_{k-1}^{\odot d}(n, t) D_1 + S_k^{\odot d}(n, t) C \right] \cdot ndt,$$

since each arriving customer chooses  $d$  servers independently and uniformly at random from the  $n$  servers, and waits for service at the server which currently contains the fewest number of customers. Similarly, the probability vector that a customer leaves a server queued by  $k$  customers in this time interval is given by

$$[-\mu S_k(n, t) + \mu S_{k+1}(n, t)] \cdot ndt.$$

Therefore, we obtain

$$\begin{aligned} dE[n_k(n, t)] &= \left[ \sum_{l=0}^{k-1} S_l^{\odot d}(n, t) D_{k-l} + S_k^{\odot d}(n, t) C \right] \cdot ndt \\ &\quad + [-\mu S_k(n, t) + \mu S_{k+1}(n, t)] \cdot ndt. \end{aligned}$$

This leads to

$$\frac{dS_k(n, t)}{dt} = \sum_{l=0}^{k-1} S_l^{\odot d}(n, t) D_{k-l} + S_k^{\odot d}(n, t) C - \mu S_k(n, t) + \mu S_{k+1}(n, t). \quad (9)$$

Since  $\lim_{n \rightarrow \infty} E[x_k^{(i)}(n, t)]$  always exists for  $k \geq 0$ , taking  $n \rightarrow \infty$  in both sides of Equation (9) we can easily obtain

$$\frac{dS_k(t)}{dt} = \sum_{l=0}^{k-1} S_l^{\odot d}(t) D_{k-l} + S_k^{\odot d}(t) C - \mu S_k(t) + \mu S_{k+1}(t). \quad (10)$$

Using a similar analysis to that in Equation (10), we obtain the system of differential vector equations for the fraction vector  $S(t) = (S_0(t), S_1(t), \dots)$  as follows:

$$S_0(t) \geq 0, S_0(t)e = 1, \quad (11)$$

$$\frac{d}{dt}S_0(t) = S_0^{\odot d}(t)C + \mu S_1(t) \quad (12)$$

and for  $k \geq 1$

$$\frac{d}{dt}S_k(t) = \sum_{l=0}^{k-1} S_l^{\odot d}(n, t) D_{k-l} + S_k^{\odot d}(n, t) C - \mu S_k(n, t) + \mu S_{k+1}(n, t). \quad (13)$$

Let  $\pi$  be the fixed point. Then  $\pi$  satisfies the following system of nonlinear equations

$$\pi_0 \geq 0, \pi_0 e = 1, \quad (14)$$

$$\pi_0^{\odot d} C + \mu \pi_1 = 0 \quad (15)$$

and for  $k \geq 1$ ,

$$\sum_{l=0}^{k-1} \pi_l^{\odot d} D_{k-l} + \pi_k^{\odot d} C - \mu \pi_k + \mu \pi_{k+1} = 0. \quad (16)$$

Let

$$Q_{\text{right}} = \begin{pmatrix} C & D_1 & D_2 & D_3 & D_4 & \cdots \\ & C & D_1 & D_2 & D_3 & \cdots \\ & & C & D_1 & D_2 & \cdots \\ & & & C & D_1 & \cdots \\ & & & & \ddots & \end{pmatrix}$$

and

$$Q_{\text{left}} = \begin{pmatrix} 0 & & & & \\ \mu I & -\mu I & & & \\ & \mu I & -\mu I & & \\ & & \mu I & -\mu I & \\ & & & \ddots & \ddots \end{pmatrix}.$$

Then the system of differential vector equations is given by

$$S_0(t) \geq 0, S_0(t)e = 1,$$

and

$$\frac{d}{dt}S(t) = S^{\odot d}(t) Q_{\text{right}} + S(t) Q_{\text{left}};$$

and the system of nonlinear equations (14) to (16) is given by

$$\pi_0 \geq 0, \pi_0 e = 1,$$

and

$$\pi^{\odot d} Q_{\text{right}} + \pi Q_{\text{left}} = 0.$$

**Remark 2** *For the supermarket model with a BMAP and exponential service times, its stochastic environment is a Markov chain of M/G/1 type whose infinitesimal generator is given by  $Q = Q_{\text{left}} + Q_{\text{right}}$ . This example clearly indicates how to set up the system of differential equations (5) and (6) for the fraction measure and the system of nonlinear equations (7) to (8) for the fixed point.*

In the remainder of this section, we provide a super-exponential solution to the fixed point  $\pi$  by means of some useful relations among the vectors  $\pi_k$  for  $k \geq 0$ .

It follows from (16) that

$$\begin{aligned} & \left( \pi_1^{\odot d}, \pi_2^{\odot d}, \pi_3^{\odot d}, \dots \right) \begin{pmatrix} C & D_1 & D_2 & \cdots \\ & C & D_1 & \cdots \\ & & C & \cdots \\ & & & \ddots \end{pmatrix} + (\pi_1, \pi_2, \pi_3, \dots) \begin{pmatrix} -\mu I & & & \\ \mu I & -\mu I & & \\ & \mu I & -\mu I & \\ & & \ddots & \ddots \end{pmatrix} \\ &= - \left( \pi_0^{\odot d} D_1, \pi_0^{\odot d} D_2, \pi_0^{\odot d} D_3, \dots \right). \end{aligned} \quad (17)$$

Let

$$A = \begin{pmatrix} -\mu I & & & \\ \mu I & -\mu I & & \\ & \mu I & -\mu I & \\ & & \ddots & \ddots \end{pmatrix}.$$

Then

$$(-A)^{-1} = \begin{pmatrix} \frac{1}{\mu} I & & & \\ \frac{1}{\mu} I & \frac{1}{\mu} I & & \\ \frac{1}{\mu} I & \frac{1}{\mu} I & \frac{1}{\mu} I & \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Note that

$$\begin{pmatrix} D_0 & D_1 & D_2 & \cdots \\ & D_0 & D_1 & \cdots \\ & & D_0 & \cdots \\ & & & \ddots \end{pmatrix} (-A^{-1}) = \begin{pmatrix} \frac{1}{\mu} \sum_{k=0}^{\infty} D_k & \frac{1}{\mu} \sum_{k=1}^{\infty} D_k & \frac{1}{\mu} \sum_{k=2}^{\infty} D_k & \cdots \\ \frac{1}{\mu} \sum_{k=0}^{\infty} D_k & \frac{1}{\mu} \sum_{k=0}^{\infty} D_k & \frac{1}{\mu} \sum_{k=1}^{\infty} D_k & \cdots \\ \frac{1}{\mu} \sum_{k=0}^{\infty} D_k & \frac{1}{\mu} \sum_{k=0}^{\infty} D_k & \frac{1}{\mu} \sum_{k=0}^{\infty} D_k & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and

$$\left( \pi_0^{\odot d} D_1, \pi_0^{\odot d} D_2, \pi_0^{\odot d} D_3, \dots \right) (-A^{-1}) = \left( \frac{1}{\mu} \pi_0^{\odot d} \sum_{k=1}^{\infty} D_k, \frac{1}{\mu} \pi_0^{\odot d} \sum_{k=2}^{\infty} D_k, \frac{1}{\mu} \pi_0^{\odot d} \sum_{k=3}^{\infty} D_k, \dots \right),$$

it follows from (17) that

$$\pi_1 = \pi_0^{\odot d} \left[ \frac{1}{\mu} \sum_{i=1}^{\infty} D_i \right] + \sum_{j=1}^{\infty} \pi_j^{\odot d} \left[ \frac{1}{\mu} \sum_{i=0}^{\infty} D_i \right] \quad (18)$$

and for  $k \geq 2$ ,

$$\pi_k = \sum_{i=0}^{k-1} \pi_i^{\odot d} \left[ \frac{1}{\mu} \sum_{j=k-i}^{\infty} D_j \right] + \sum_{j=k}^{\infty} \pi_j^{\odot d} \left[ \frac{1}{\mu} \sum_{i=0}^{\infty} D_i \right]. \quad (19)$$

To omit the terms  $\sum_{j=k}^{\infty} \pi_j^{\odot d} \left[ \frac{1}{\mu} \sum_{i=0}^{\infty} D_i \right]$  for  $k \geq 1$ , we assume that the system of nonlinear equations (18) and (19) has a closed-form solution

$$\pi_k = r(k) \gamma^{\odot \frac{1}{d}}, \quad (20)$$

where  $r(k)$  is an underdetermined positive constant for  $k \geq 1$ . Then it follows from (18), (19) and (20) that

$$\pi_1 = \pi_0^{\odot d} \left[ \frac{1}{\mu} \sum_{i=1}^{\infty} D_i \right] \quad (21)$$

or

$$r(1) \gamma^{\odot \frac{1}{d}} = \pi_0^{\odot d} \left[ \frac{1}{\mu} \sum_{i=1}^{\infty} D_i \right]; \quad (22)$$

and for  $k \geq 2$ ,

$$\pi_k = \sum_{i=0}^{k-1} \pi_i^{\odot d} \left[ \frac{1}{\mu} \sum_{j=k-i}^{\infty} D_j \right]$$

or

$$r(k) \gamma^{\odot \frac{1}{d}} = \pi_0^{\odot d} \left[ \frac{1}{\mu} \sum_{j=k}^{\infty} D_j \right] + \sum_{i=0}^{k-1} [r(i)]^d \gamma \left[ \frac{1}{\mu} \sum_{j=k-i}^{\infty} D_j \right]. \quad (23)$$

Let  $\theta = 1/\gamma^{\odot \frac{1}{d}} e$ . Then  $0 < \theta < 1$ . Let  $\lambda_k = \gamma \sum_{i=k}^{\infty} D_i e$  and  $\rho_k = \lambda_k/\mu$ . Then it follows from (22) and (23) that

$$r(1) = \frac{\theta}{\mu} \pi_0^{\odot d} \sum_{i=1}^{\infty} D_i e \quad (24)$$

and for  $k \geq 2$

$$\begin{aligned} r(k) &= \frac{\theta}{\mu} \pi_0^{\odot d} \sum_{j=k}^{\infty} D_j e + \frac{\theta}{\mu} \sum_{i=1}^{k-1} [r(i)]^d \gamma \sum_{j=k-i}^{\infty} D_j e \\ &= \frac{\theta}{\mu} \pi_0^{\odot d} \sum_{j=k}^{\infty} D_j e + \theta \sum_{i=1}^{k-1} [r(i)]^d \rho_{k-i}. \end{aligned} \quad (25)$$

It is easy to see from (24) and (25) that  $\pi_0$  and  $r(1)$  are two key underdetermined terms for the closed-form solution to the system of nonlinear equations (22) and (23). Let us first derive the vector  $\pi_0$ . It follows from (15) and (21) that

$$\begin{cases} \pi_0^{\odot d} C + \mu \pi_1 = 0, \\ \mu \pi_1 = \pi_0^{\odot d} \sum_{i=1}^{\infty} D_i. \end{cases}$$

This leads to

$$\pi_0^{\odot d} \left( C + \sum_{i=1}^{\infty} D_i \right) = 0.$$

Thus, it is easy to see that  $\pi_0 = \theta \gamma^{\odot \frac{1}{d}}$ , which is a probability vector with  $\pi_0 e = 1$ . Hence we have

$$\pi_1 = -\frac{\theta^d}{\mu} \gamma C = \frac{\theta^d}{\mu} \gamma \sum_{i=1}^{\infty} D_i. \quad (26)$$

It follows from (24) and (25) that

$$r(1) = \frac{\theta}{\mu} \cdot \theta^d \gamma \sum_{i=1}^{\infty} D_i e = \theta^{d+1} \rho_1 \quad (27)$$

and for  $k \geq 2$

$$\begin{aligned} r(k) &= \frac{\theta}{\mu} \pi_0^{\odot d} \sum_{j=k}^{\infty} D_j e + \frac{\theta}{\mu} \sum_{i=1}^{k-1} [r(i)]^d \gamma \sum_{j=k-i}^{\infty} D_j e \\ &= \theta^{d+1} \rho_k + \theta \sum_{i=1}^{k-1} [r(i)]^d \rho_{k-i}. \end{aligned} \quad (28)$$

Therefore, we obtain the super-exponential solution to the fixed point as follows:

$$\pi_0 = \theta \gamma^{\odot \frac{1}{d}}$$

and for  $k \geq 1$

$$\pi_k = \left[ \theta^{d+1} \rho_k + \theta \sum_{i=1}^{k-1} [r(i)]^d \rho_{k-i} \right] \gamma^{\odot \frac{1}{d}}.$$

## 4 A Supermarket Model of GI/M/1 Type

In this section, we analyze a supermarket model with Poisson arrivals and batch PH service processes. Note that the stochastic environment is a Markov chain of GI/M/1 type, thus the supermarket model is called to be of GI/M/1 type. For the supermarket model of GI/M/1 type, we set up the system of differential equations for the fraction measure by means of density-dependent jump Markov processes, and derive the system of nonlinear equations satisfied the fixed point which can be computed by an iterative algorithm. Further, it is seen that the supermarket model of GI/M/1 type is more difficult than the case of M/G/1 type.

Let us describe the supermarket model of GI/M/1 type. Customers arrive at a queueing system of  $n > 1$  servers as a Poisson process with arrival rate  $n\lambda$  for  $\lambda > 0$ . The service times of each batch of customers are of phase type with irreducible representation  $(\alpha, T)$  of order  $m$  and with a batch size distribution  $\{b_k, k = 1, 2, 3, \dots\}$  for  $\sum_{k=1}^{\infty} b_k = 1$  and  $\bar{b} = \sum_{k=1}^{\infty} kb_k < +\infty$ . Let  $T^0 = -Te \not\geq 0$ . Then the expected service time is given by  $1/\mu = -\alpha T^{-1}e = \eta T^0$ , where  $\eta$  is the stationary probability vector of the Markov chain  $T + T^0\alpha$ . Each batch of arriving customers choose  $d \geq 1$  servers independently and uniformly at random from the  $n$  servers, and waits for service at the server which currently contains the fewest number of customers. If there is a tie, servers with the fewest number of customers will be chosen randomly. All customers in every server will be served in FCFS for different batches and in random service within one batch. We assume that all random variables defined above are independent of each other, and that the system is operating in the stable region  $\rho = \lambda/\mu\bar{b} < 1$ . Clearly,  $d$  is an input choice number in this supermarket model. Figure 2 is depicted as an illustration for supermarket models of GI/M/1 type.

We define  $n_k^{(i)}(t)$  as the number of queues with at least  $k$  customers and the service time in phase  $i$  at time  $t \geq 0$ . Clearly,  $0 \leq n_k^{(i)}(t) \leq n$  for  $k \geq 1$  and  $1 \leq i \leq m$ . Let

$$X_n^{(0)}(t) = \frac{n}{n} = 1,$$



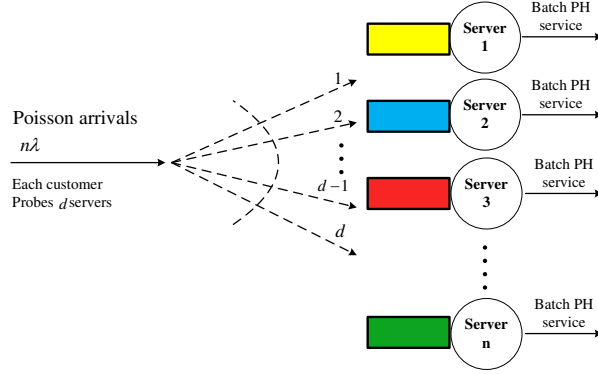


Figure 2: A supermarket model of GI/M/1 type

and  $k \geq 1$

$$X_n^{(k,i)}(t) = \frac{n_k^{(i)}(t)}{n},$$

which is the fraction of queues with at least  $k$  customers and the service time in phase  $i$  at time  $t \geq 0$ . We write

$$X_n^{(k)}(t) = \left( X_n^{(k,1)}(t), X_n^{(k,2)}(t), \dots, X_n^{(k,m)}(t) \right), \quad k \geq 1,$$

$$X_n(t) = \left( X_n^{(0)}(t), X_n^{(1)}(t), X_n^{(2)}(t), \dots \right).$$

The state of the supermarket model may be described by the vector  $X_n(t)$  for  $t \geq 0$ . Since the arrival process to the queueing system is Poisson and the service times of each server are of phase type,  $\{X_n(t), t \geq 0\}$  is a Markov process whose state space is given by

$$\Omega_n = \left\{ \left( g_n^{(0)}, g_n^{(1)}, g_n^{(2)}, \dots \right) : g_n^{(0)} = 1, g_n^{(k-1)} \geq g_n^{(k)} \geq 0, \right. \\ \left. \text{and } ng_n^{(k)} \text{ is a vector of nonnegative integers for } k \geq 1 \right\}.$$

Let

$$s_0(n, t) = E \left[ X_n^{(0)}(t) \right]$$

and  $k \geq 1$

$$s_k^{(i)}(n, t) = E \left[ X_n^{(k,i)}(t) \right].$$

Clearly,  $s_0(n, t) = 1$ . We write

$$S_k(n, t) = \left( s_k^{(1)}(n, t), s_k^{(2)}(n, t), \dots, s_k^{(m)}(n, t) \right), \quad k \geq 1.$$

As shown in Martin and Suhov [18] and Luczak and McDiarmid [14], the Markov process  $\{X_n(t), t \geq 0\}$  is asymptotically deterministic as  $n \rightarrow \infty$ . Thus  $\lim_{n \rightarrow \infty} E[X_n^{(0)}(t)]$  and  $\lim_{n \rightarrow \infty} E[X_n^{(k,i)}]$  always exist by means of the law of large numbers. Based on this, we write

$$S_0(t) = \lim_{n \rightarrow \infty} s_0(n, t) = 1,$$

for  $k \geq 1$

$$s_k^{(i)}(t) = \lim_{n \rightarrow \infty} s_k^{(i)}(n, t),$$

$$S_k(t) = \left(s_k^{(1)}(t), s_k^{(2)}(t), \dots, s_k^{(m)}(t)\right)$$

and

$$S(t) = (S_0(t), S_1(t), S_2(t), \dots).$$

Let  $X(t) = \lim_{n \rightarrow \infty} X_n(t)$ . Then it is easy to see from Poisson arrivals and batch PH service times that  $\{X(t), t \geq 0\}$  is also a Markov process whose state space is given by

$$\Omega = \left\{ \left( g^{(0)}, g^{(1)}, g^{(2)}, \dots \right) : g^{(0)} = 1, g^{(k-1)} \geq g^{(k)} \geq 0 \right\}.$$

If the initial distribution of the Markov process  $\{X_n(t), t \geq 0\}$  approaches the Dirac delta-measure concentrated at a point  $g \in \Omega$ , then its steady-state distribution is concentrated in the limit on the trajectory  $S_g = \{S(t) : t \geq 0\}$ . This indicates a law of large numbers for the time evolution of the fraction of queues of different lengths. Furthermore, the Markov process  $\{X_n(t), t \geq 0\}$  converges weakly to the fraction vector  $S(t) = (S_0(t), S_1(t), S_2(t), \dots)$ , or for a sufficiently small  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P\{\|X_n(t) - S(t)\| \geq \varepsilon\} = 0,$$

where  $\|a\|$  is the  $L_\infty$ -norm of vector  $a$ .

To determine the fraction vector  $S(t)$ , we need to set up a system of differential vector equations satisfied by the fraction measure  $S(t)$  by means of density-dependent jump Markov processes. Consider the supermarket model with  $n$  servers, and determine the expected change in the number of queues with at least  $k$  customers over a small time period of length  $dt$ . The probability vector that during this time period, any arriving customer joins a queue of size  $k - 1$  is given by

$$n \left[ \lambda S_{k-1}^{\odot d}(n, t) - \lambda S_k^{\odot d}(n, t) \right] dt.$$

Similarly, the probability vector that a customer leaves a server queued by  $k$  customers is given by

$$n \left[ S_k(n, t) T + \sum_{l=1}^{\infty} b_l S_{k+l}(n, t) T^0 \alpha \right] dt.$$

Therefore, we obtain

$$\begin{aligned} dE[n_k(n, t)] = & n \left[ \lambda S_{k-1}^{\odot d}(n, t) - \lambda S_k^{\odot d}(n, t) \right] dt \\ & + n \left[ S_k(n, t) T + \sum_{l=1}^{\infty} b_l S_{k+l}(n, t) T^0 \alpha \right] dt. \end{aligned}$$

This leads to

$$\frac{dS_k(n, t)}{dt} = \lambda S_{k-1}^{\odot d}(n, t) - \lambda S_k^{\odot d}(n, t) + S_k(n, t) T + \sum_{l=1}^{\infty} b_l S_{k+l}(n, t) T^0 \alpha. \quad (29)$$

Taking  $n \rightarrow \infty$  in both sides of Equation (29), we have

$$\frac{dS_k(t)}{dt} = \lambda S_{k-1}^{\odot d}(t) - \lambda S_k^{\odot d}(t) + S_k(t) T + \sum_{l=1}^{\infty} b_l S_{k+l}(t) T^0 \alpha. \quad (30)$$

Using a similar analysis to Equation (30), we obtain a system of differential vector equations for the fraction vector  $S(t) = (S_0(t), S_1(t), S_2(t), \dots)$  as follows:

$$S_0(t) = 1, \quad (31)$$

$$\frac{d}{dt} S_0(t) = -\lambda S_0^d(t) + \sum_{l=1}^{\infty} S_l(t) T^0 \sum_{k=l}^{\infty} b_k, \quad (32)$$

$$\frac{d}{dt} S_1(t) = \lambda \alpha S_0^d(t) - \lambda S_1^{\odot d}(t) + S_1(t) T + \sum_{l=1}^{\infty} b_l S_{1+l}(t) T^0 \alpha, \quad (33)$$

and for  $k \geq 2$ ,

$$\frac{d}{dt} S_k(t) = \lambda S_{k-1}^{\odot d}(t) - \lambda S_k^{\odot d}(t) + S_k(t) T + \sum_{l=1}^{\infty} b_l S_{k+l}(t) T^0 \alpha. \quad (34)$$

If the row vector  $\pi = (\pi_0, \pi_1, \pi_2, \dots)$  is a fixed point of the fraction vector  $S(t)$ , then the fixed point  $\pi$  satisfies the following system of nonlinear equations

$$\pi_0 = 1 \quad (35)$$

$$-\lambda \pi_0^d + \sum_{l=1}^{\infty} \pi_l T^0 \sum_{k=l}^{\infty} b_k = 0, \quad (36)$$

$$\lambda\alpha\pi_0^d - \lambda\pi_1^{\odot d} + \pi_1 T + \sum_{l=1}^{\infty} b_l \pi_{1+l} T^0 \alpha = 0, \quad (37)$$

and for  $k \geq 2$ ,

$$\lambda\pi_{k-1}^{\odot d} - \lambda\pi_k^{\odot d} + \pi_k T + \sum_{l=1}^{\infty} b_l \pi_{k+l} T^0 \alpha = 0. \quad (38)$$

Let

$$Q_{\text{right}} = \begin{pmatrix} -\lambda & \lambda\alpha & & & \cdots \\ & -\lambda I & \lambda I & & \cdots \\ & & -\lambda I & \lambda I & \cdots \\ & & & -\lambda I & \lambda I & \cdots \\ & & & & \ddots & \ddots \end{pmatrix}$$

and

$$Q_{\text{left}} = \begin{pmatrix} 0 & & & & & \\ T^0 & T & & & & \\ T^0 \sum_{k=2}^{\infty} b_k & b_1 T^0 \alpha & T & & & \\ T^0 \sum_{k=3}^{\infty} b_k & b_2 T^0 \alpha & b_1 T^0 \alpha & T & & \\ T^0 \sum_{k=4}^{\infty} b_k & b_3 T^0 \alpha & b_2 T^0 \alpha & b_1 T^0 \alpha & T & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Then the system of differential vector equations for the fraction measure is given by

$$S_0(t) = 1,$$

and

$$\frac{d}{dt} S(t) = S^{\odot d}(t) Q_{\text{right}} + S(t) Q_{\text{left}};$$

and the system of nonlinear equations for the fixed point is given by

$$\pi_0 = 1,$$

and

$$\pi^{\odot d} Q_{\text{right}} + \pi Q_{\text{left}} = 0. \quad (39)$$

In the remainder of this section, we provide an iterative algorithm for computing the fixed point for the supermarket model of GI/M/1 type. Specifically, the iterative algorithm indicates that the supermarket model of GI/M/1 type is more difficult than the case of M/G/1 type.

Let

$$B = \begin{pmatrix} -\lambda I & \lambda I & & \cdots \\ & -\lambda I & \lambda I & \cdots \\ & & -\lambda I & \cdots \\ & & & \ddots \end{pmatrix}$$

and

$$Q_{\text{service}} = \begin{pmatrix} T & & & & \\ b_1 T^0 \alpha & T & & & \\ b_2 T^0 \alpha & b_1 T^0 \alpha & T & & \\ b_3 T^0 \alpha & b_2 T^0 \alpha & b_1 T^0 \alpha & T & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Then it follows from (39) that

$$\pi_0^d(\lambda\alpha, 0, 0, 0, \dots) + \pi_{\mathcal{L}}^{\odot d} B + \pi_{\mathcal{L}} Q_{\text{service}} = 0, \quad (40)$$

where  $\pi_{\mathcal{L}} = (\pi_1, \pi_2, \pi_3, \dots)$ . Note that

$$(-B)^{-1} = \begin{pmatrix} \frac{1}{\lambda} I & \frac{1}{\lambda} I & \frac{1}{\lambda} I & \cdots \\ & \frac{1}{\lambda} I & \frac{1}{\lambda} I & \cdots \\ & & \frac{1}{\lambda} I & \cdots \\ & & & \ddots \end{pmatrix},$$

using  $\pi_0 = 1$  we obtain

$$\pi_0^d(\lambda\alpha, 0, 0, 0, \dots) (-B)^{-1} = (\alpha, \alpha, \alpha, \dots)$$

and

$$Q_{\text{service}} (-B)^{-1} = \frac{1}{\lambda} \begin{pmatrix} T & T & T & T & \cdots \\ (T^0 \alpha) b_1 & T + (T^0 \alpha) b_1 & T + (T^0 \alpha) b_1 & T + (T^0 \alpha) b_1 & \cdots \\ (T^0 \alpha) b_2 & (T^0 \alpha) \sum_{k=1}^2 b_k & T + (T^0 \alpha) \sum_{k=1}^2 b_k & T + (T^0 \alpha) \sum_{k=1}^2 b_k & \cdots \\ (T^0 \alpha) b_3 & (T^0 \alpha) \sum_{k=2}^3 b_k & (T^0 \alpha) \sum_{k=1}^3 b_k & T + (T^0 \alpha) \sum_{k=1}^3 b_k & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

Thus it follows from (40) that

$$\begin{aligned} \pi_{\mathcal{L}}^{\odot d} &= \pi_0^d(\lambda\alpha, 0, 0, 0, \dots) (-B)^{-1} + \pi_{\mathcal{L}} Q_{\text{service}} (-B)^{-1} \\ &= (\alpha, \alpha, \alpha, \dots) + \pi_{\mathcal{L}} Q_{\text{service}} (-B)^{-1}. \end{aligned}$$

This leads to that for  $k \geq 1$

$$\lambda \pi_k^d = \lambda \alpha + \sum_{l=1}^{k+1} \pi_l T + b_1 \sum_{l=2}^{k+2} \pi_l (T^0 \alpha) + b_2 \sum_{l=3}^{k+3} \pi_l (T^0 \alpha) + \dots. \quad (41)$$

To solve the system of nonlinear equations (41) for  $k \geq 1$ , we assume that the fixed point has a closed-form solution

$$\pi_k = r(k) \eta,$$

where  $\eta$  is the stationary probability vector of the Markov chain  $T + T^0 \alpha$ . It follows from (41) that

$$\lambda r^d(k) \eta^{\odot d} = \lambda \alpha + \sum_{l=1}^{k+1} r(l) \eta T + b_1 \sum_{l=2}^{k+2} r(l) \eta (T^0 \alpha) + b_2 \sum_{l=3}^{k+3} r(l) \eta (T^0 \alpha) + \dots.$$

Taking  $\theta = \eta^{\odot d} e$ . Then  $\theta \in (0, 1)$ . Noting that  $\alpha e = 1$ ,  $\eta T e = -\mu$  and  $\eta T^0 = \mu$ , we obtain that for  $k \geq 1$

$$\rho \theta r^d(k) = \rho - \sum_{l=1}^{k+1} r(l) + b_1 \sum_{l=2}^{k+2} r(l) + b_2 \sum_{l=3}^{k+3} r(l) + \dots. \quad (42)$$

This gives

$$\rho \theta \left[ r^d(k) - r^d(k+1) \right] = r(k+2) - \sum_{l=1}^{\infty} b_l r(k+2+l). \quad (43)$$

Thus it follows from (43) that

$$\rho \theta \left( r^d(1) - r^d(2), r^d(2) - r^d(3), r^d(3) - r^d(4), \dots \right) = (r(3), r(4), r(5), \dots) C, \quad (44)$$

where

$$C = \begin{pmatrix} 1 & & & & \\ -b_1 & 1 & & & \\ -b_2 & -b_1 & 1 & & \\ -b_3 & -b_2 & -b_1 & 1 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Therefore, we have

$$C^{-1} = \begin{pmatrix} 1 & & & & \\ b_1 & & 1 & & \\ b_2 + b_1^2 & & b_1 & & 1 \\ b_3 + 2b_2b_1 + b_1^3 & & b_2 + b_1^2 & & b_1 & 1 \\ b_4 + 2b_3b_1 + 3b_2b_1^2 + b_1^4 & & b_3 + 2b_2b_1 + b_1^3 & & b_2 + b_1^2 & b_1 & 1 \\ \vdots & & \vdots & & \vdots & \vdots & \ddots \end{pmatrix},$$

and the norm  $\|\cdot\|_\infty$  of the matrix  $C^{-1}$  is given by

$$\|C^{-1}\| = \sup_{i \geq j \geq 1} \{|\zeta_{i,j}|\} \leq \sum_{k=1}^{\infty} b_k = 1,$$

where  $\zeta_{i,j}$  is the  $(i,j)$ th entry of the matrix  $C^{-1}$  for  $0 \leq j \leq i$ . Note that the norm  $\|C^{-1}\| \leq 1$  is useful for our following iterative algorithm designed by the matrix  $\rho\theta C^{-1}$  with  $\|\rho\theta C^{-1}\| = \rho\theta < 1$ .

Let

$$X = (r^d(1) - r^d(2), r^d(2) - r^d(3), r^d(3) - r^d(4), \dots)$$

and

$$Y = (r(3), r(4), r(5), \dots).$$

Then

$$X = (r^d(1), r^d(2), Y^{\odot d}) - (r^d(2), Y^{\odot d})$$

and it follows from (44) that

$$\begin{aligned} Y &= X(\rho\theta C^{-1}) \\ &= (r^d(1), r^d(2), Y^{\odot d})(\rho\theta C^{-1}) - (r^d(2), Y^{\odot d})(\rho\theta C^{-1}). \end{aligned} \quad (45)$$

It follows from (42) that

$$r(1) = \rho - \rho\theta r^d(1) - r(2)(1 - b_1) + Y(b_1 + b_2, b_2 + b_3, b_3 + b_4, \dots)^T \quad (46)$$

and

$$r(2) = \rho - r(1) - [b_1 r(2) + \rho\theta r^d(2)] + Y(b_1 + b_2 - 1, b_1 + b_2 + b_3, b_2 + b_3 + b_4, \dots)^T \quad (47)$$

Now, we use Equations (45) to (47) to provide an iterative algorithm for computing the fixed point  $\pi_k = r(k)\eta$  for  $k \geq 1$ . To that end, we write

$$Y_N = (r_N(3), r_N(4), r_N(5), \dots) \quad (48)$$

and

$$R_N = (r_N(1), r_N(2), r_N(3), \dots) = (r_N(1), r_N(2), Y_N). \quad (49)$$

Let

$$\begin{aligned} r_{N+1}(1) = & \rho - \rho \theta r_N^d(1) - r_N(2)(1 - b_1) \\ & + Y_N(b_1 + b_2, b_2 + b_3, b_3 + b_4, \dots)^T, \end{aligned} \quad (50)$$

$$\begin{aligned} r_{N+1}(2) = & \rho - r_N(1) - [b_1 r_N(2) + \rho \theta r_N^d(2)] \\ & + Y_N(b_1 + b_2 - 1, b_1 + b_2 + b_3, b_2 + b_3 + b_4, \dots)^T \end{aligned} \quad (51)$$

and

$$Y_{N+1} = \left( r_N^d(1), r_N^d(2), Y_N^{\odot d} \right) (\rho \theta C^{-1}) - \left( r_N^d(2), Y_N^{\odot d} \right) (\rho \theta C^{-1}). \quad (52)$$

Based on the iterative relations given in (50) to (52), we provide an iterative algorithm for computing the vector  $R = (r(1), r(2), r(3), \dots)$ . This gives the fixed point  $\pi = (1, r(1)\eta, r(2)\eta, r(3)\eta, \dots)$ .

**An Iterative Algorithm:** Computation of the Fixed Point

**Input:**  $\lambda, (\alpha, T), \{b_k\}$  and  $d$ .

**Output:**  $R = (r(1), r(2), r(3), \dots)$  and  $\pi = (1, r(1)\eta, r(2)\eta, r(3)\eta, \dots)$ .

**Computational Steps:**

*Step one:* Taking the initial value  $R_0 = 0$ , that is,  $r_0(1) = 0, r_0(2) = 0, Y_0 = 0$ .

*Step two:* Computing  $R_1 = (r_1(1), r_1(2), Y_1)$  through

$$r_1(1) = \rho, \quad \leftarrow (50)$$

$$r_1(2) = \rho, \quad \leftarrow (51)$$

$$Y_1 = (\rho^d, \rho^d, 0) (\rho \theta C^{-1}) - (\rho^d, 0) (\rho \theta C^{-1}), \quad \leftarrow (52)$$

*Step three:* If  $R_N$  is known, computing  $R_{N+1} = (r_{N+1}(1), r_{N+1}(2), Y_{N+1})$  through

$$\begin{aligned} r_{N+1}(1) = & \rho - \rho \theta r_N^d(1) - r_N(2)(1 - b_1) \\ & + Y_N(b_1 + b_2, b_2 + b_3, b_3 + b_4, \dots)^T, \end{aligned} \quad \leftarrow (50)$$

$$\begin{aligned} r_{N+1}(2) = & \rho - r_N(1) - [b_1 r_N(2) + \rho \theta r_N^d(2)] \\ & + Y_N(b_1 + b_2 - 1, b_1 + b_2 + b_3, b_2 + b_3 + b_4, \dots)^T, \end{aligned} \quad \leftarrow (51)$$

$$Y_{N+1} = \left( r_N^d(1), r_N^d(2), Y_N^{\odot d} \right) (\rho \theta C^{-1}) - \left( r_N^d(2), Y_N^{\odot d} \right) (\rho \theta C^{-1}). \quad \leftarrow (52)$$

*Step four:* For a sufficiently small  $\varepsilon > 0$ , if there exists Step  $K$  such that  $\|R_{K+1} - R_K\| < \varepsilon$ , then our computation is end in this step; otherwise we go to Step three for continuous computations.



*Step five:* When our computation is over at Step  $K$ , computing

$$\pi = (1, r_K(1)\eta, r_K(2)\eta, r_K(3)\eta, \dots)$$

as an approximate fixed point under an error  $\varepsilon > 0$ .

In what follows we analyze two numerical examples by means of the above iterative algorithm.

In the first example, we take

$$\lambda = 1, d = 2, \alpha = (1/2, 1/2),$$

$$T(1) = \begin{pmatrix} -4 & 3 \\ 2 & -7 \end{pmatrix}, T(2) = \begin{pmatrix} -5 & 3 \\ 2 & -7 \end{pmatrix}, T(3) = \begin{pmatrix} -4 & 4 \\ 2 & -7 \end{pmatrix},$$

Table 1 illustrates how the super-exponential solution ( $\pi_1$  to  $\pi_5$ ) depends on the matrices  $T(1)$ ,  $T(2)$  and  $T(3)$ , respectively.

Table 1: The super-exponential solution depends on the matrix  $T$

|         | $T(1)$                 | $T(2)$                 | $T(3)$                 |
|---------|------------------------|------------------------|------------------------|
| $\pi_1$ | (0.2045, 0.1591)       | (0.1410, 0.1026)       | (0.3125, 0.2500)       |
| $\pi_2$ | (0.0137, 0.0107)       | (0.0043, 0.0031)       | (0.0500, 0.0400)       |
| $\pi_3$ | (6.193e-05, 4.817e-05) | (3.965e-06, 2.884e-06) | (0.0013, 0.0010)       |
| $\pi_4$ | (1.259e-09, 9.793e-10) | (3.390e-12, 2.465e-12) | (8.446e-07, 6.757e-07) |
| $\pi_5$ | (5.204e-19, 4.048e-19) | (2.478e-24, 1.802e-24) | (3.656e-13, 2.925e-13) |

In the second example, we take

$$\lambda = 1, d = 5, \alpha(1) = (1/3, 1/3, 1/3), \alpha(2) = (1/12, 7/12, 1/3),$$

$$T = \begin{pmatrix} -10 & 2 & 4 \\ 3 & -7 & 4 \\ 0 & 2 & -5 \end{pmatrix},$$

Table 2 shows how the super-exponential solution ( $\pi_1$  to  $\pi_4$ ) depends on the vectors  $\alpha(1)$  and  $\alpha(2)$ , respectively.

## 5 Supermarket Models with Multiple Choices

In this section, we consider two supermarket models with multiple choices: The first one is one mobile server with multiple waiting lines under the service discipline of joint-shortest

Table 2: The super-exponential solution depends on the vectors  $\alpha$

|         | $\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ | $\alpha = (\frac{1}{12}, \frac{7}{12}, \frac{1}{3})$ |
|---------|--|--|
| $\pi_1$ | (0.0741, 0.1358 , 0.2346)                          | (0.0602, 0.1728, 0.2531)                             |
| $\pi_2$ | (5.619e-05, 1.030e-05, 1.779e-04 )                 | (7.182e-05, 2.063e-04, 3.020e-04)                    |
| $\pi_3$ | (1.411e-20, 2.587e-20, 4.469e-20)                  | (1.739e-19, 4.993e-19, 7.311e-19)                    |
| $\pi_4$ | (1.410e-98, 2.586e-98, 4.466e-98)                  | (1.444e-92, 4.148e-92, 6.074e-92)                    |

queue and serve-longest queue, and the second one is a supermarket model with multiple classes of Poisson arrivals, each of which has a choice number. Our main purpose is to organize the system of nonlinear equations for the fixed point under multiple choice numbers, and to be able to obtain super-exponential solution to the fixed points for the two supermarket models.

### 5.1 One mobile server with multiple waiting lines

The supermarket model is structured as one mobile server with multiple waiting lines, where the Poisson arrivals joint a waiting line with the shortest queue and the mobile server enters a waiting line with the longest queue for his service woks. Such a system is depicted in Figure 3 for an illustration. For one mobile server with  $n$  waiting lines, customers arrive at this system as a Poisson process with arrival rate  $n\lambda$ , and all customers are served by one mobile server with service rate  $n\mu$ . Each arriving customer chooses  $d \geq 1$  waiting lines independently and uniformly at random from the  $n$  waiting lines, and waits for service at a waiting line which currently contains the fewest number of customers. If there is a tie, waiting lines with the fewest number of customers will be chosen by the arriving customer randomly. The mobile server chooses  $f \geq 1$  waiting lines independently and uniformly at random from the  $n$  waiting lines, and enters a waiting line which currently contains the most number of customers. If there is a tie, waiting lines with the most number of customers will be chosen be the server randomly. All customers in every waiting line will be served in the FCFS manner. We assume that all random variables defined above are independent of each other, and that the system is operating in the stable region  $\rho = \lambda/\mu < 1$ . Clearly,  $d$  and  $f$  are input choice number and output choice number in this supermarket model, respectively.

It is clear that the stochastic environment of this supermarket model is a positive recurrent birth-death process with an irreducible infinitesimal generator  $Q = Q_{\text{left}} + Q_{\text{right}}$ ,

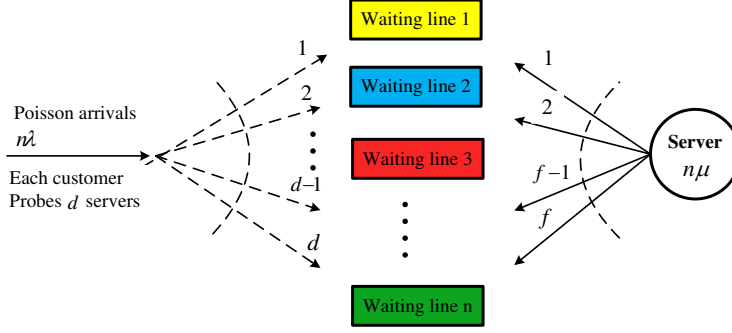


Figure 3: A supermarket model with input and output choices

where

$$Q_{\text{left}} = \begin{pmatrix} 0 & & & & \\ \mu & -\mu & & & \\ & \mu & -\mu & & \\ & & \mu & -\mu & \\ & & & \ddots & \ddots \end{pmatrix}$$

and

$$Q_{\text{right}} = \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & -\lambda & \lambda & \\ & & & \ddots & \ddots \end{pmatrix}.$$

Similar derivation to those given in Section 3 or 4, we obtain that the fixed point satisfies the system of nonlinear equations

$$\pi_0 = 1$$

and

$$\pi^{\odot f} Q_{\text{left}} + \pi^{\odot d} Q_{\text{right}} = 0. \quad (53)$$

Let

$$Q = \begin{pmatrix} Q_{0,0} & U \\ V & Q^{(\mathcal{L})} \end{pmatrix},$$

where

$$Q_{0,0} = -\lambda, U = (\lambda, 0, 0, \dots), V = (\mu, 0, 0, \dots)^T,$$

$$Q_{\text{arrival}}^{(\mathcal{L})} = \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & -\lambda & \lambda & \\ & & & \ddots & \ddots \\ & & & & \ddots & \ddots \end{pmatrix}$$

and

$$Q_{\text{service}}^{(\mathcal{L})} = \begin{pmatrix} -\mu & & & & \\ \mu & -\mu & & & \\ & \mu & -\mu & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \end{pmatrix}.$$

It follows from (53) that

$$\pi_0 = 1, \quad (54)$$

$$-\lambda\pi_0^d + \mu\pi_1^f = 0 \quad (55)$$

and

$$\pi_0^d U + \pi_{\mathcal{L}}^{\odot d} Q_{\text{arrival}}^{(\mathcal{L})} + \pi_{\mathcal{L}}^{\odot f} Q_{\text{service}}^{(\mathcal{L})} = 0. \quad (56)$$

It follows from (54) and (55) that

$$\pi_1 = \rho^{\frac{1}{f}}.$$

Note that

$$\left[-Q_{\text{service}}^{(\mathcal{L})}\right]^{-1} = \begin{pmatrix} \frac{1}{\mu} & & & & \\ \frac{1}{\mu} & \frac{1}{\mu} & & & \\ \frac{1}{\mu} & \frac{1}{\mu} & \frac{1}{\mu} & & \\ \vdots & \vdots & \vdots & \ddots & \end{pmatrix},$$

it follows from (56) that for  $k \geq 2$

$$\pi_k^f = \pi_{k-1}^d \rho.$$

This leads to

$$\pi_k = \rho^{\frac{\sum_{i=0}^{k-1} d^i f^{k-1-i}}{f^k}} = \rho^{\frac{1}{f} \sum_{i=0}^{k-1} \left(\frac{d}{f}\right)^i}. \quad (57)$$

Specifically, when  $d \neq f$ , we have

$$\pi_k = \rho^{\frac{\left(\frac{d}{f}\right)^k - 1}{d-f}}.$$

**Remark 3** Equation (57) indicates different influence of the input and output choice numbers  $d$  and  $f$  on the fixed point  $\pi$ . If  $d > f$ , then the fixed point  $\pi$  decreases doubly exponentially; and if  $d = f$ , then  $\pi_k = \rho^{\frac{k}{f}}$  which is geometric. However, it is very interesting for the case with  $d < f$ . In this case,  $\lim_{k \rightarrow \infty} \pi_k = \rho^{\frac{1}{f-d}}$ , which illustrates that the fraction of waiting lines with infinite customers has a positive lower bound  $\rho^{\frac{1}{f-d}} > 0$ . This shows that if  $\rho < 1$  and  $d < f$ , this supermarket model is transient.

## 5.2 A supermarket model with multiple input choices

Now, we analyze a supermarket model with multiple input choices. There are  $m$  types of different customers who arrive at a queueing system of  $n > 1$  servers for receiving their required service. Arrivals of customers of  $i$ th type are a Poisson process with arrival rate  $n\lambda_i$  for  $\lambda_i > 0$ , and the service times at each server are exponential with service rate  $\mu > 0$ . Note that different types of customers have the same service time. Each arriving customer of  $i$ th type chooses  $d_i \geq 1$  servers independently and uniformly at random from the  $n$  servers, and waits for service at the server which currently contains the fewest number of customers. If there is a tie, servers with the fewest number of customers will be chosen randomly. All customers in every server will be served in the FCFS manner. We assume that all random variables defined above are independent of each other, and that the system is operating in the stable region  $\rho = \sum_{i=1}^m \rho_i < 1$ , where  $\rho_i = \lambda_i/\mu$ . Clearly,  $d_1, d_2, \dots, d_m$  are multiple input choice numbers in this supermarket model.

Let

$$Q_{\text{right}}(i) = \begin{pmatrix} -\lambda_i & \lambda_i & & & \\ & -\lambda_i & \lambda_i & & \\ & & -\lambda_i & \lambda_i & \\ & & & \ddots & \ddots \end{pmatrix}$$

and

$$Q_{\text{left}} = \begin{pmatrix} 0 & & & & \\ \mu & -\mu & & & \\ & \mu & -\mu & & \\ & & & \ddots & \ddots \end{pmatrix}.$$

Obviously, the stochastic environment of this supermarket model is a positive recurrent birth-death process with an irreducible infinitesimal generator  $Q = Q_{\text{left}} + \sum_{i=1}^m Q_{\text{right}}(i)$ .

Similar derivation to those given in Section 3 or 4, we obtain that the fixed point satisfies the system of nonlinear equations

$$\pi_0 = 1 \quad (58)$$

and

$$\pi Q_{\text{left}} + \sum_{i=1}^m \pi^{\odot d_i} Q_{\text{right}}(i) = 0. \quad (59)$$

Let

$$\begin{aligned} \pi &= (\pi_0, \pi_{\mathcal{L}}), \\ U_i &= (\lambda_i, 0, 0, \dots), \\ Q_{\text{arrival}}^{(\mathcal{L})}(i) &= \begin{pmatrix} -\lambda_i & \lambda_i & & & \\ & -\lambda_i & \lambda_i & & \\ & & -\lambda_i & \lambda_i & \\ & & & \ddots & \ddots \end{pmatrix} \end{aligned}$$

and

$$Q_{\text{service}}^{(\mathcal{L})} = \begin{pmatrix} -\mu & & & & \\ \mu & -\mu & & & \\ & \mu & -\mu & & \\ & & & \ddots & \ddots \end{pmatrix}.$$

Therefore, the system of nonlinear equations (58) and (59) is written as

$$\pi_0 = 1, \quad (60)$$

$$-\sum_{i=1}^m \lambda_i \pi_0^{d_i} + \mu \pi_1 = 0, \quad (61)$$

$$\sum_{i=1}^m \pi_0^{d_i} (\lambda_i, 0, 0, \dots) + \sum_{i=1}^m \pi_{\mathcal{L}}^{\odot d_i} Q_{\text{arrival}}^{(\mathcal{L})}(i) + \pi_{\mathcal{L}} Q_{\text{service}}^{(\mathcal{L})} = 0. \quad (62)$$

It follows from (60) and (61) that

$$\pi_1 = \sum_{i=1}^m \rho_i = \rho,$$

and from (62) that

$$\pi_{\mathcal{L}} = \sum_{i=1}^m \pi_0^{d_i} (\lambda_i, 0, 0, \dots) \left[ -Q_{\text{service}}^{(\mathcal{L})} \right]^{-1} + \sum_{i=1}^m \pi_{\mathcal{L}}^{\odot d_i} Q_{\text{arrival}}^{(\mathcal{L})}(i) \left[ -Q_{\text{service}}^{(\mathcal{L})} \right]^{-1}.$$

This leads to that for  $k \geq 2$

$$\pi_k = \sum_{i=1}^m \pi_{k-1}^{d_i} \rho_i.$$

Let  $\delta_1 = \rho$  and  $\delta_k = \sum_{i=1}^m \delta_{k-1}^{d_i} \rho_i$  for  $k \geq 2$ . Then the fixed point has a super-exponential solution

$$\pi_0 = 1$$

and for  $k \geq 1$

$$\pi_k = \delta_k.$$

## Acknowledgements

Q.L. Li was supported by the National Science Foundation of China under grant No. 10871114.

## References

- [1] Y. Azar, A.Z. Broder, A.R. Karlin and E. Upfal (1999). Balanced allocations. *SIAM Journal on Computing* **29**, 180–200.
- [2] M. Bramson, Y. Lu and B. Prabhakar (2010). Randomized load balancing with general service time distributions. In *Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 275–286.
- [3] M. Dahlin (1999). Interpreting stale load information. *IEEE Transactions on Parallel and Distributed Systems* **11**, 1033–1047.
- [4] C. Graham (2000). Kinetic limits for large communication networks. In *Modelling in Applied Sciences*, N. Bellomo and M. Pulvirenti (eds.), Birkhäuser, pages 317–370.
- [5] C. Graham (2000). Chaoticity on path space for a queueing network with selection of the shortest queue among several. *Journal of Applied Probability* **37**, 198–201.
- [6] C. Graham (2004). Functional central limit theorems for a large network in which customers join the shortest of several queues. *Probability Theory Related Fields* **131**, 97–120.

- [7] M. Harchol-Balter and A.B. Downey (1997). Exploiting process lifetime distributions for dynamic load balancing. *ACM Transactions on Computer Systems* **15**, 253–285.
- [8] T.G. Kurtz (1981). *Approximation of Population Processes*. SIAM.
- [9] Q.L. Li (2010). *Constructive Computation in Stochastic Models with Applications: The RG-Factorizations*. Springer and Tsinghua Press.
- [10] Q.L. Li (2010). Doubly exponential solution for randomized load balancing with general service times. Submitted for publication.
- [11] Q.L. Li and John C.S. Lui (2010). Doubly exponential solution for randomized load balancing models with Markovian arrival processes and PH service times. Submitted for publication.
- [12] Q.L. Li, John C.S. Lui and Y. Wang (2010). A matrix-analytic solution for randomized load balancing models with PH service times. In *Proceeding of PERFORM 2010 Workshop*, Lecture Notes of Computer Science, Pages 1-20.
- [13] Q.L. Li, John C.S. Lui and Y. Wang (2010). Super-exponential solution for a retrieval supermarket Mmodel. Submitted for publication.
- [14] M. Luczak and C. McDiarmid (2006). On the maximum queue length in the supermarket model. *The Annals of Probability* **34**, 493–527.
- [15] M. Luczak and C. McDiarmid (2007). Asymptotic distributions and chaos for the supermarket model. *Electronic Journal of Probability* **12**, 75–99.
- [16] M.J. Luczak and J.R. Norris (2005). Strong approximation for the supermarket model. *The Annals of Applied Probability* **15**, 2038–2061.
- [17] J.B. Martin (2001). Point processes in fast Jackson networks. *The Annals of Applied Probability* **11**, 650-663.
- [18] J.B. Martin and Y.M. Suhov (1999). Fast Jackson networks. *The Annals of Applied Probability* **9**, 854–870.
- [19] M.D. Mitzenmacher (1996). The power of two choices in randomized load balancing. PhD thesis, University of California at Berkeley, Department of Computer Science, Berkeley, CA.



- [20] M. Mitzenmacher (1999). On the analysis of randomized load balancing schemes. *Theory of Computing Systems* **32**, 361–386.
- [21] M. Mitzenmacher (2000). How useful is old information? *IEEE Transactions on Parallel and Distributed Systems* **11**, 6–20.
- [22] M. Mitzenmacher, A. Richa, and R. Sitaraman (2001). The power of two random choices: a survey of techniques and results. In *Handbook of Randomized Computing: Volume 1*, P. Pardalos, S. Rajasekaran and J. Rolim (eds), pages 255–312.
- [23] R. Mirchandaney, D. Towsley, and J.A. Stankovic (1989). Analysis of the effects of delays on load sharing. *IEEE Transactions on Computers* **38**, 1513–1525.
- [24] M.F. Neuts (1989). *Structured stochastic matrices of M/G/1 type and their applications*. Marcel Dekker Inc.: New York.
- [25] Y.M. Suhov and N.D. Vvedenskaya (2002). Fast Jackson Networks with Dynamic Routing. *Problems of Information Transmission* **38**, 136{153.
- [26] B. Vöcking (1999). How asymmetry helps load balancing. In *Proceedings of the Fortieth Annual Symposium on Foundations of Computer Science*, pages 131–140.
- [27] N.D. Vvedenskaya, R.L. Dobrushin and F.I. Karpelevich (1996). Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problems of Information Transmissions* **32**, 20–34.
- [28] N.D. Vvedenskaya and Y.M. Suhov (1997). Dobrushin’s mean-field approximation for a queue with dynamic routing. *Markov Processes and Related Fields* **3**, 493–526.
- [29] R. Weber (1978). On the optimal assignment of customers to parallel servers. *Journal of Applied Probabilities* **15**, 406–413.
- [30] W. Winston (1977). Optimality of the shortest line discipline. *Journal of Applied Probabilities* **14**, 181–189.